Reviewer's report

Title: Doublet method for very fast autocoding

Version: Date: 11 August 2004

Reviewer: Christopher G Chute

Reviewer's report:

This is an interesting and intriguing representation of a fast indexing technique optimized for medical concepts. The notion of doublet coding is well recognized in the computational linguistic literature, which does not appear to be referenced. Indeed, most of that literature attempts to identify optimal combinations of terms balancing concept disambiguation with combinatorial explosions. Specifically, the literature suggests that while doublets are good, sometimes triplets or four-plexes are even better.

The author does not address the question of word order dependence within the phrases that are being indexed. Specifically, for very long terms which might include seven or eight words, their specific order will determine the exact doublets that are created. Variation in that internal word order, which is often quite arbitrary, will profoundly influence the exact doublets that would subsequently be created for indexing. The traditional solution to this is to create all possible doublets within a phrase; an approach does that does not appear to be mentioned by the author.

There are significant stylistic problems with the manuscript in its present form. Grammatical and usage errors are striking.

While the elucidation of software functionality through the use of code snippets is perhaps appropriate for a computer monthly, a more detailed description of the underlying algorithms and their optimization might be of more interest to most readers. While this reviewer could follow the Perl extracts with some facility, this is not likely to be the case for the majority of readers.

Overall, the presentation of doublet coding is both intriguing and convincing. However, modest refinements of this presentation would make the work that much more generalizable.

What next?: Accept after minor essential revisions

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: No

Declaration of competing interests:

None